

(19)日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開2002-49395

(P2002-49395A)

(43)公開日 平成14年2月15日(2002.2.15)

(51)IntCl. <sup>7</sup>	識別記号	F I	キーワード(参考)
G 1 0 L 19/04		H 0 3 M 7/36	5 D 0 4 5
11/04		G 1 0 L 101:12	5 J 0 6 4
H 0 3 M 7/36		9/18	D
// G 1 0 L 101:12		9/08	B
		9/14	J

審査請求 未請求 請求項の数 8 O L (全 10 頁)

(21)出願番号 特願2000-238891(P2000-238891)

(22)出願日 平成12年8月2日(2000.8.2)

(71)出願人 000002185

ソニー株式会社

東京都品川区北品川6丁目7番35号

(72)発明者 近藤 哲二郎

東京都品川区北品川6丁目7番35号ソニー株式会社内

(72)発明者 渡辺 勉

東京都品川区北品川6丁目7番35号ソニー株式会社内

(74)代理人 100082740

弁理士 田辺 恵基

Fターム(参考) 5D045 CC10 DA02

5J064 AA01 AA02 BA01 BA06 BB03

BB04 BB10 BC01 BC27 BD03

(54)【発明の名称】 デジタル信号処理方法、学習方法及びそれらの装置並びにプログラム格納媒体

(57)【要約】

【課題】デジタル信号の波形再現性を一段と向上し得るデジタル信号処理方法、学習方法及びそれらの装置並びにプログラム格納媒体を提案する。

【解決手段】デジタルオーディオ信号D10のピッチを検出し、当該検出したピッチに基づいてそのクラスを分類し、分類されたクラスに対応した予測方式でデジタルオーディオ信号D10を変換するようにしたことにより、一段とデジタルオーディオ信号D10の特徴に適応した変換を行うことができる。

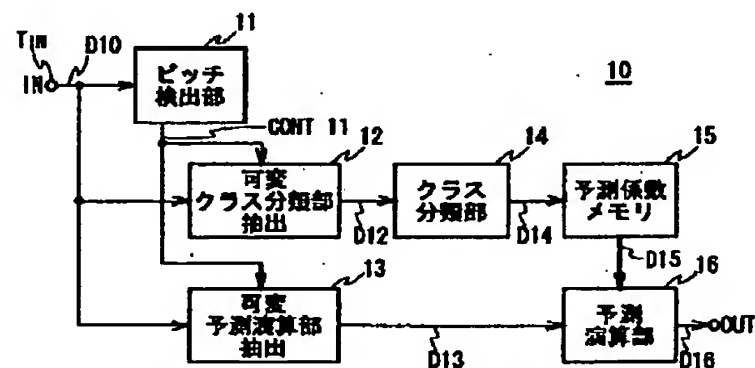


図1 オーディオ信号処理装置の構成

## 【特許請求の範囲】

【請求項1】 デジタルオーディオ信号を変換するデジタル信号処理装置において、

上記デジタルオーディオ信号のピッチを検出するピッチ検出手段と、

上記ピッチに基づいてそのクラスを分類するクラス分類手段と、

上記分類されたクラスに対応した予測方式で上記デジタルオーディオ信号を予測演算することにより上記デジタルオーディオ信号を変換してなる新たなデジタルオーディオ信号を生成する予測演算手段とを具えることを特徴とするデジタル信号処理装置。

【請求項2】 上記予測演算手段は、予め所望とするデジタルオーディオ信号に基づいて学習により生成されている予測係数を用いることを特徴とする請求項1に記載のデジタル信号処理装置。

【請求項3】 デジタルオーディオ信号を変換するデジタル信号処理方法において、

上記デジタルオーディオ信号のピッチを検出するピッチ検出ステップと、

上記ピッチに基づいてそのクラスを分類するクラス分類ステップと、

上記分類されたクラスに対応した予測方式で上記デジタルオーディオ信号を予測演算することにより上記デジタルオーディオ信号を変換してなる新たなデジタルオーディオ信号を生成する予測演算ステップとを具えることを特徴とするデジタル信号処理方法。

【請求項4】 上記予測演算ステップでは、予め所望とするデジタルオーディオ信号に基づいて学習により生成されている予測係数が用いられることを特徴とする請求項3に記載のデジタル信号処理方法。

【請求項5】 デジタルオーディオ信号を変換するデジタル信号処理装置の上記変換処理の予測演算に用いられる予測係数を生成する学習装置において、

所望とするデジタルオーディオ信号から当該デジタルオーディオ信号を劣化させた生徒デジタルオーディオ信号を生成する生徒デジタルオーディオ信号生成手段と、

上記生徒デジタルオーディオ信号のピッチを検出するピッチ検出手段と、

上記ピッチに基づいてそのクラスを分類するクラス分類手段と、

上記デジタルオーディオ信号と上記生徒デジタルオーディオ信号とに基づいて上記クラスに対応する予測係数を算出する予測係数算出手段とを具えることを特徴とする学習装置。

【請求項6】 デジタルオーディオ信号を変換するデジタル信号処理装置の上記変換処理の予測演算に用いられる予測係数を生成する学習方法において、

所望とするデジタルオーディオ信号から当該ディ

ジタルオーディオ信号を劣化させた生徒デジタルオーディオ信号を生成する生徒デジタルオーディオ信号生成ステップと、

上記生徒デジタルオーディオ信号のピッチを検出するピッチ検出ステップと、

上記ピッチに基づいてそのクラスを分類するクラス分類ステップと、

上記デジタルオーディオ信号と上記生徒デジタルオーディオ信号とに基づいて上記クラスに対応する予測係数を算出する予測係数算出ステップとを具えることを特徴とする学習方法。

【請求項7】 上記デジタルオーディオ信号のピッチを検出するピッチ検出ステップと、

上記ピッチに基づいてそのクラスを分類するクラス分類ステップと、

上記分類されたクラスに対応した予測方式で上記デジタルオーディオ信号を予測演算することにより上記デジタルオーディオ信号を変換してなる新たなデジタルオーディオ信号を生成する予測演算ステップとを含むプログラムをデジタル信号処理装置に実行させるプログラム格納媒体。

【請求項8】 所望とするデジタルオーディオ信号から当該デジタルオーディオ信号を劣化させた生徒デジタルオーディオ信号を生成する生徒デジタルオーディオ信号生成ステップと、

上記生徒デジタルオーディオ信号のピッチを検出するピッチ検出ステップと、

上記ピッチに基づいてそのクラスを分類するクラス分類ステップと、

上記デジタルオーディオ信号と上記生徒デジタルオーディオ信号とに基づいて上記クラスに対応する予測係数を算出する予測係数算出ステップとを含むプログラムを学習装置に実行させるプログラム格納媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明はデジタル信号処理方法、学習方法及びそれらの装置並びにプログラム格納媒体に関し、レートコンバータ又はPCM(Pulse Code Modulation) 復号装置等においてデジタル信号に対してデータの補間処理を行うデジタル信号処理方法、学習方法及びそれらの装置並びにプログラム格納媒体に適用して好適なものである。

【0002】

【従来の技術】 従来、デジタルオーディオ信号をデジタル／アナログコンバータに入力する前に、サンプリング周波数を元の値の数倍に変換するオーバーサンプリング処理を行っている。これにより、デジタル／アナログコンバータから出力されたデジタルオーディオ信号はアナログ・アンチ・エイリアス・フィルタの位相特性が可聴周波数高域で一定に保たれ、また、サンプリング

に伴うデジタル系のイメージ雑音の影響が排除されるようになされている。

【0003】かかるオーバーサンプリング処理では、通常、線形一次（直線）補間方式のデジタルフィルタが用いられている。このようなデジタルフィルタは、サンプリングレートが変わったりデータが欠落した場合等に、複数の既存データの平均値を求めて直線的な補間データを生成するものである。

【0004】

【発明が解決しようとする課題】ところが、オーバーサンプリング処理後のデジタルオーディオ信号は、線形一次補間によって時間軸方向に対してデータ量が数倍に緻密になっているものの、オーバーサンプリング処理後のデジタルオーディオ信号の周波数帯域は変換前とあまり変わらず、音質そのものは向上していない。さらに、補間されたデータは必ずしもA/D変換前のアナログオーディオ信号の波形に基づいて生成されたのではないため、波形再現性もほとんど向上していない。

【0005】また、サンプリング周波数の異なるデジタルオーディオ信号をダビングする場合において、サンプリング・レート・コンバータを用いて周波数を変換しているが、かかる場合でも線形一次デジタルフィルタによって直線的なデータの補間しか行うことができず、音質や波形再現性を向上することが困難であった。さらに、デジタルオーディオ信号のデータサンプルが欠落した場合において同様である。

【0006】本発明は以上の点を考慮してなされたもので、デジタル信号の波形再現性を一段と向上し得るデジタル信号処理方法、学習方法及びそれらの装置並びにプログラム格納媒体を提案しようとするものである。

【0007】

【課題を解決するための手段】かかる課題を解決するため本発明においては、デジタルオーディオ信号のピッチを検出し、当該検出したピッチに基づいてそのクラスを分類し、分類されたクラスに対応した予測方式でデジタルオーディオ信号を変換するようにしたことにより、一段とデジタルオーディオ信号の特徴に適応した変換を行うことができる。

【0008】

【発明の実施の形態】以下図面について、本発明の一実施の形態を詳述する。

【0009】図1においてオーディオ信号処理装置10は、デジタルオーディオ信号（以下これをオーディオデータと呼ぶ）のサンプリングレートを上げたり、オーディオデータを補間する際に、真値に近いオーディオデータをクラス分類適用処理によって生成するようになされている。因みに、デジタルオーディオ信号とは、人や動物が発する声を表す音声信号、楽器が発する楽音を表す楽音信号、及びその他の音を表す信号を意味するものである。

【0010】すなわち、オーディオ信号処理装置10において、ピッチ検出部11は入力端子T<sub>IN</sub>から供給された図2又は図3に示す入力オーディオデータD10のピッチ（基本周波数）を例えば入力オーディオデータD10の自己相関に基づいて検出する。

【0011】すなわちピッチ検出部11は、入力オーディオデータD10について、例えば時間軸領域に分割してなる種々の範囲ごとに自己相関値を検出し、当該自己相関値が所定の閾値よりも大きいときピッチが有ると判断する。そしてピッチ検出部11はピッチが有る場合のピッチ周期PIT<sub>i</sub>及びピッチサイズとその位置をピッチクラスとする。例えば、入力オーディオデータD10が図2に示すような波形を有している場合、ピッチ検出部11はカレントピッチPIT<sub>2</sub>に対してその前後に隣接する位置に前ピッチPIT<sub>1</sub>及び後ろピッチPIT<sub>3</sub>を検出する。因みに、ピッチ検出部11が検出する前ピッチPIT<sub>1</sub>及び後ろピッチPIT<sub>3</sub>は必ずしもカレントピッチPIT<sub>2</sub>に隣接するとは限らず、カレントピッチPIT<sub>2</sub>から所定距離離れている場合もある。

【0012】ピッチ検出部11はカレントピッチPIT<sub>2</sub>に対応する前ピッチPIT<sub>1</sub>及び後ろピッチPIT<sub>3</sub>を検出すると、これらを切り出すための切り出し制御信号CONT11を可変クラス分類部抽出部12及び可変予測演算部抽出部13に供給する。

【0013】可変クラス分類部抽出部12は切り出し制御信号CONT11に基づいて、前ピッチPIT<sub>1</sub>に対応するクラスタップCL1、カレントピッチPIT<sub>2</sub>に対応するクラスタップCL2及び後ろタップPIT<sub>3</sub>に対応するクラスタップCL3を入力オーディオデータD10から切り出し、これらをクラスタップデータD12としてクラス分類部14に供給する。

【0014】クラス分類部14は、クラス分類抽出部12において切り出されたクラスタップについて、各ピッチ（前ピッチPIT<sub>1</sub>、カレントピッチPIT<sub>2</sub>及び後ろタップPIT<sub>3</sub>）ごとに個別に、又はこれらを統合し、そのクラスタップを圧縮して圧縮データパターンを生成するADRC(Adaptive Dynamic Range Coding)回路部と、クラスタップデータD12の属するクラスコードを発生するクラスコード発生回路部とを有する。

【0015】ADRC回路部はクラスタップデータD12に対して、例えば8ビットから2ビットに圧縮するような演算を行うことによりパターン圧縮データを形成する。このADRC回路部は、適応的量子化を行うものであり、ここでは、信号レベルの局所的なパターンを短い語長で効率的に表現することができるので、信号パターンのクラス分類のコード発生用に用いられる。

【0016】具体的には、オーディオ波形上の6つの8ビットのデータ（クラスタップ）をクラス分類しようとする場合、2<sup>48</sup>という膨大な数のクラスに分類しなければならず、回路上の負担が多くなる。そこで、この実施

の形態のクラス分類部14ではその内部に設けられたADRRC回路部で生成されるパターン圧縮データに基づいてクラス分類を行う。例えば6つのクラスタップに対して1ビットの量子化を実行すると、6つのクラスタップを6ビットで表すことができ、 $2^6 = 64$ クラスに分類することができる。

$$DR = MAX - MIN + 1$$

$$Q = \{ (L - MIN + 0.5) \times 2^m / DR \} \quad \dots\dots (1.)$$

【0019】に従って、領域内の最大値MAXと最小値MINとの間を指定されたビット長で均等に分割して量子化を行う。なお、(1)式において{ }は小数点以下の切り捨て処理を意味する。かくして6つのクラスタップが、それぞれ例えば8ビット( $m=8$ )で構成されているとすると、これらはADRRC回路部においてそれぞれが2ビットに圧縮される。

$$class = \sum_{i=1}^n q_i (2^P)^i \quad \dots\dots (2)$$

【0022】に示す演算を実行することにより、そのブロック( $q_1 \sim q_n$ )が属するクラスを示すクラスコードclassを算出すると共に、当該算出されたクラスタップデータD12に基づくクラスコードclassをクラスコードデータD14として予測係数メモリ15に供給する。このクラスコードclassは、予測係数メモリ15から予測係数を読み出す際の読み出しアドレスを示す。因みに(2)式において、nは圧縮されたオーディオ波形データ(クラスタップ) $q_i$ の数を表し、この実施の形態の場合 $n=6$ であり、またPはビット割り当てを表し、この実施の形態の場合 $P=2$ である。

【0023】このようにして、クラス分類部14はクラス分類部抽出部12において入力オーディオデータD10から切り出されたクラスタップのクラスコードclassを生成し、これをクラスコードデータD14として予測★

$$y' = w_1 x_1 + w_2 x_2 + \dots\dots + w_n x_n \quad \dots\dots (3)$$

【0027】に示す積和演算を行うことにより、予測結果 $y'$ を得る。この予測値 $y'$ が、音質が改善されたオーディオデータD16として予測演算部16から出力される。

【0028】因みに、ピッチ検出部11において入力オーディオデータD10にピッチが無いと判断した場合、ピッチ検出部11は前ピッチPIT1や後ろピッチPIT3を用いることなく、図3に示すようにカレントデータ近傍のオーディオ波形のレベルを切り出すための制御信号CONT11を可変クラス分類部抽出部12及び可変予測演算部抽出部13に供給することにより、カレントデータ近傍のオーディオ波形によりクラス分類及び予測演算を行う。

【0029】なお、オーディオ信号処理装置10の構成

\*【0017】ここで、ADRRC回路部は、切り出された領域内のクラスタップ(オーディオ波形)のダイナミックレンジをDR、ビット割り当てをm、各クラスタップのデータレベルをL、量子化コードをQとすると、次式、

$$【0018】$$

$$* 【数1】$$

\*【0020】このようにして圧縮されたオーディオ波形データ(クラスタップ)をそれぞれ $q_n$  ( $n=1 \sim 6$ )とすると、クラス分類部14に設けられたクラスコード発生回路部は、圧縮されたオーディオ波形データ $q_n$ に基づいて、次式、

$$【0021】$$

$$* 【数2】$$

20★係数メモリ15に供給する。

【0024】予測係数メモリ15には、各クラスコードに対応する予測係数のセットがクラスコードに対応するアドレスにそれぞれ記憶されており、クラス分類部14から供給されるクラスコードデータD14に基づいて、当該クラスコードに対応するアドレスに記憶されている予測係数のセット $w_1 \sim w_n$ が読み出され、予測演算部16に供給される。

【0025】予測演算部16は、予測演算部抽出部13においてピッチ検出部11から供給される切り出し制御信号CONT11に応じて切り出された予測演算しようとするオーディオ波形データ(予測タップ)D13( $x_1 \sim x_n$ )と、予測係数 $w_1 \sim w_n$ に対して、次式

$$【0026】$$

$$【数3】$$

$$\dots\dots (3)$$

として図1について上述した機能ブロックを示したが、この機能ブロックを構成する具体的構成として、この実施の形態においては図4に示すコンピュータ構成の装置を用いる。すなわち、図4において、オーディオ信号処理装置10は、バスBUSを介してCPU21、ROM(Read Only Memory)22、予測係数メモリ15を構成するRAM(Random Access Memory)15、及び各回路部がそれぞれ接続された構成を有し、CPU11はROM22に格納されている種々のプログラムを実行することにより、図1について上述した各機能ブロック(ピッチ検出部11、クラス分類部抽出部12、予測演算部抽出部13、クラス分類部14及び予測演算部16)として動作するようになされている。

50 【0030】また、オーディオ信号処理装置10にはネ

ットワークとの間で通信を行う通信インターフェース24、フロッピーディスクや光磁気ディスク等の外部記憶媒体から情報を読み出すリムーバブルドライブ28を有し、ネットワーク経由又は外部記憶媒体から図1について上述したクラス分類適用処理を行うための各プログラムをハードディスク装置25のハードディスクに読み込み、当該読み込まれたプログラムに従ってクラス分類適用処理を行うこともできる。

【0031】ユーザは、キーボードやマウス等の入力手段26を介して種々のコマンドを入力することにより、CPU21に対して図1について上述したクラス分類処理を実行させる。この場合、オーディオ信号処理装置10はデータ入出力部27を介して音質を向上させようとするオーディオデータ（入力オーディオデータ）D10を入力し、当該入力オーディオデータD10に対してクラス分類適用処理を施した後、音質が向上したオーディオデータD16をデータ入出力部27を介して外部に出力し得るようになされている。

【0032】因みに、図5はオーディオ信号処理装置10におけるクラス分類適用処理の処理手順を示し、オーディオ信号処理装置10はステップSP11から当該処理手順に入ると、続くステップSP12において入力オーディオデータD10のピッチをピッチ検出部11において算出する。

【0033】この算出されたピッチ（前ピッチP1T1、カレントピッチP1T2、後ろピッチP1T3）はオーディオ波形のクラス分類を一段と確実にするためのものであり、オーディオ信号処理装置10は、ステップSP13においてピッチに応じたクラスタップの切り出しを行った後、ステップSP14においてクラス分類部14によりクラスタップ（オーディオ波形）をクラス分類する。そしてオーディオ信号処理装置10は、クラス分類の結果得られたクラスコードを用いて予測係数メモリ15から予測係数を読み出す。この予測係数は予め学習によりクラス毎に対応して格納されており、オーディオ信号処理装置10はクラスコードに対応した予測係数を読み出すことにより、このときのオーディオ波形の特徴に合致した予測係数を用いることができる。

【0034】予測係数メモリ15から読み出された予測係数は、ステップSP15において予測演算部16の予測演算に用いられる。これにより、入力オーディオデータD10はそのピッチに応じたクラスタップにより得られたクラスコードを用いて予測演算され、所望とするオーディオデータD16に変換される。かくして入力オーディオデータD10はその音質が改善されたオーディオデータD16に変換され、オーディオ信号処理装置10はステップSP16に移って当該処理手順を終了する。

【0035】次に、図1について上述した予測係数メモリ15に記憶するクラス毎の予測係数のセットを予め学習によって得るための学習回路について説明する。

【0036】図6において、学習回路30は、高音質の教師オーディオデータD30を生徒信号生成フィルタ37に受ける。生徒信号生成フィルタ37は、間引き率設定信号D39により設定された間引き率で教師オーディオデータD30を所定時間ごとに所定サンプル間引くようになされている。

【0037】この場合、生徒信号生成フィルタ37における間引き率によって、生成される予測係数が異なり、これに応じて上述のオーディオ信号処理装置10で再現されるオーディオデータも異なる。例えば、上述のオーディオ信号処理装置10においてサンプリング周波数を高くすることでオーディオデータの音質を向上しようとする場合、生徒信号生成フィルタ37ではサンプリング周波数を減らす間引き処理を行う。また、これに対して上述のオーディオ信号処理装置10において入力オーディオデータD10の欠落したデータサンプルを補うことで音質の向上を図る場合には、これに応じて、生徒信号生成フィルタ37ではデータサンプルを欠落させる間引き処理を行うようになされている。

【0038】かくして、生徒信号生成フィルタ37は教師オーディオデータ30から所定の間引き処理により生徒オーディオデータD37を生成し、これをピッチ検出部31、クラス分類部抽出部32及び予測演算部抽出部33にそれぞれ供給する。

【0039】ピッチ検出部31は生徒信号生成フィルタ37から供給された生徒オーディオデータD37について、図2及び図3について上述したピッチ（基本周波数）を生徒オーディオデータD37の自己相関に基づいて検出する。

【0040】そしてピッチ検出部31はカレントピッチP1T2に対応する前ピッチP1T1及び後ろピッチP1T3を検出すると、これらを切り出すための切り出し制御信号CONT31を可変クラス分類部抽出部32及び可変予測演算部抽出部33に供給する。

【0041】可変クラス分類部抽出部32は切り出し制御信号CONT31に基づいて、前ピッチP1T1に対応するクラスタップCL1、カレントピッチP1T2に対応するクラスタップCL2及び後ろタップP1T3に対応するクラスタップCL3を生徒オーディオデータD37から切り出し、これらをクラスタップデータD32としてクラス分類部34に供給する。

【0042】クラス分類部34は、クラス分類抽出部32において切り出されたクラスタップについて、当該クラスタップを圧縮して圧縮データパターンを生成するADRC(Adaptive Dynamic Range Coding) 回路部と、クラスタップデータD32の属するクラスコードを発生するクラスコード発生回路部とを有する。

【0043】ADRC回路部はクラスタップデータD32に対して、例えば8ビットから2ビットに圧縮するような演算を行うことによりパターン圧縮データを形成す



る。このADRC回路部は、適応的量子化を行うものであり、ここでは、信号レベルの局所的なパターンを短い語長で効率的に表現することができるので、信号パターンのクラス分類のコード発生用に用いられる。

【0044】具体的には、包絡線波形上の6つの8ビットのデータ（クラスタップ）をクラス分類しようとする場合、 $2^{48}$ という膨大な数のクラスに分類しなければならず、回路上の負担が多くなる。そこで、この実施の形態のクラス分類部14ではその内部に設けられたADRC回路部で生成されるパターン圧縮データに基づいてクラスタップの量子化を実行すると、6つのクラスタップを6ビットで表すことができ、 $2^6 = 64$ クラスに分類することができる。

【0045】ここで、ADRC回路部は、切り出された領域内のクラスタップ（オーディオ波形）のダイナミックレンジをDR、ビット割り当をm、各クラスタップのデータレベルをL、量子化コードをQとして、上述の

(1)式と同様の演算により、領域内の最大値MAXと最小値MINとの間を指定されたビット長で均等に分割して量子化を行う。かくしてオーディオ波形上の6つの波形データが、それぞれ例えば8ビット（ $m=8$ ）で構成されているとすると、これらはADRC回路部においてそれぞれが2ビットに圧縮される。

【0046】このようにして圧縮されたオーディオ波形データ（クラスタップ）をそれぞれ $q_i$ （ $i=1\sim6$ ）とすると、クラス分類部34に設けられたクラスコード発生回路部は、圧縮されたオーディオ波形データ $q_i$ に基づいて、上述の(2)式と同様の演算を実行することにより、そのブロック（ $q_1\sim q_6$ ）が属するクラスを示すクラスコードclassを算出し、当該算出されたクラ

$$y = w_1 x_1 + w_2 x_2 + \dots + w_n x_n \quad \dots (4)$$

【0051】とする。学習前は、 $w_n$ が未定係数である。

【0052】学習回路30では、クラスコード毎に、複数のオーディオデータに対して学習を行う。データサン

$$y_k = w_1 x_{k1} + w_2 x_{k2} + \dots + w_n x_{kn} \quad \dots (5)$$

【0054】が設定される。但し $k=1, 2, \dots, M$ である。

【0055】 $M>n$ の場合、予測係数 $w_1, \dots, w_n$ は一意的に決まらないので、誤差ベクトル $e$ の要素を次

$$e_k = y_k - \{w_1 x_{k1} + w_2 x_{k2} + \dots + w_n x_{kn}\} \quad \dots (6)$$

【0057】によって定義し（但し、 $k=1, 2, \dots, M$ ）、次式、

$$e^2 = \sum_{k=1}^M e_k^2 \quad \dots (7)$$

【0059】を最小にする予測係数を求める。いわゆる、最小自乗法による解法である。

【0060】ここで、(7)式による $w_n$ の偏微分係数

\* スコードclassをクラスコードデータD34として予測係数算出部36に供給する。因みに(2)式において、 $n$ は圧縮されたオーディオ波形データ（クラスタップ） $q_i$ の数を表し、この実施の形態の場合 $n=6$ であり、また $P$ はビット割り当てを表し、この実施の形態の場合 $P=2$ である。

【0047】このようにして、クラス分類部34はクラスコードデータD34を生成し、これを予測係数算出部36に供給する。また、予測係数算出部36には、予測演算部抽出部33においてピッチ検出部11から供給される切り出し制御信号CONT11に応じて切り出された予測演算しようとするオーディオ波形データ（予測タ

ップ）D13（ $x_1 \sim x_n$ ）が供給される。

【0048】予測係数算出部36は、クラス分類部34から供給されたクラスコードclassと、各クラスコードclass毎に切り出された予測タップD33と、入力端T<sub>in</sub>から供給された高音質の教師オーディオデータD30とを用いて、正規方程式を立てる。

【0049】すなわち、生徒オーディオデータD37の $n$ サンプルのレベルをそれぞれ $x_1, x_2, \dots, x_n$ として、それぞれに $p$ ビットのADRCを行った結果の量子化データを $q_1, \dots, q_n$ とする。このとき、この領域のクラスコードclass'を上述の(2)式のように定義する。そして、上述のように生徒オーディオデータD37のレベルをそれぞれ、 $x_1, x_2, \dots, x_n$ とし、高音質の教師オーディオデータD30のレベルを $y$ としたとき、クラスコード毎に、予測係数 $w_1, w_2, \dots, w_n$ による $n$ タップの線形推定式を設定する。これを次式、

【0050】

【数4】

\* ブル数がMの場合、上述の(4)式に従って、次式、

【0053】

【数5】

★式、

【0056】

【数6】

☆【0058】

☆【数7】

を求める。この場合、次式、

【0061】

【数8】

$$\frac{\partial e^2}{\partial w_i} = \sum_{k=1}^n 2 \left[ \frac{\partial e_k}{\partial w_i} \right] e_k = \sum_{k=1}^n 2 x_{ki} \cdot e_k$$

$$= \sum_{k=1}^n 2 x_{ki} \cdot e_k \quad (i=1, 2, \dots, n) \quad \dots (8)$$

【0062】を「0」にするように、各 $w_n$  ( $n=1 \sim 6$ )を求めれば良い。 \* 【0064】  
【数9】

【0063】そして、次式、 \*

$$X_{ij} = \sum_{p=0}^M x_{pi} \cdot x_{pj} \quad \dots (9)$$

【0065】 ※ ※ 【数10】

$$Y_i = \sum_{k=1}^n x_{ki} \cdot y_k \quad \dots (10)$$

【0066】のように、 $X_{ij}$ 、 $Y_i$ を定義すると、

(8)式は行列を用いて次式、

★ 【0067】

★ 【数11】

$$\begin{bmatrix} X_{11} & X_{12} & \dots & X_{1n} \\ X_{21} & X_{22} & \dots & X_{2n} \\ \vdots & \vdots & & \vdots \\ X_{m1} & X_{m2} & \dots & X_{mn} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \quad \dots (11)$$

【0068】として表される。

【0069】この方程式は、一般に正規方程式と呼ばれている。なお、ここでは $n=6$ である。

【0070】全ての学習用データ(教師オーディオデータD30、クラスコードclass、予測タップD33)の入力が完了した後、予測係数算出部36は各クラスコードclassに上述の(11)式に示した正規方程式を立て、この正規方程式を掃き出し法等の一般的な行列解法を用いて、各 $w_n$ について解き、各クラスコード毎に、予測係数を算出する。予測係数算出部36は、算出された各予測係数(D36)を予測係数メモリ15に書き込む。

【0071】このような学習を行った結果、予測係数メモリ15には、量子化データ $q_1, \dots, q_6$ で規定されるパターン毎に、高音質のオーディオデータ $y$ を推定するための予測係数が、各クラスコード毎に格納される。この予測係数メモリ15は、図1について上述したオーディオ信号処理装置10において用いられる。かかる処理により、線形推定式に従って通常のオーディオデータから高音質のオーディオデータを作成するための予測係数の学習が終了する。

【0072】このように、学習回路30は、オーディオ信号処理装置10において補間処理を行う程度を考慮して、生徒信号生成フィルタ37で高音質の教師オーディオデータの間引き処理を行うことにより、オーディオ信号処理装置10における補間処理のための予測係数を生成することができる。

【0073】以上の構成において、オーディオ信号処理装置10は、入力オーディオデータD10のピッチをピッチ検出部11において検出することにより、入力オーディオデータD10のうち、特に規則性を持った波形部分を抽出することができる。

【0074】かかる規則性を持った波形部分(ピッチ)は、例えば人間の声では同じ音素であっても個人毎に異なり、この部分を正確に切り出してクラス分類することにより、個人毎に異なるオーディオ波形の特徴が一段と明確にクラス分けされる。

【0075】このように、オーディオ波形の特徴を明確にクラス分けし、当該クラス分類結果に基づく予測演算を行うことにより、オーディオ波形の再現性が向上し、音質の向上が図られる。

【0076】以上の構成によれば、ピッチの検出結果に基づいてタップの切り出し範囲を制御するようにしたことにより、入力オーディオデータD10を一段と高音質のオーディオデータD16に変換することができる。

【0077】なお上述の実施の形態においては、オーディオ信号処理装置10及び学習回路30において、ピッチ検出部11、31で検出されたピッチサイズとその位置とをピッチクラスとして可変クラス分類部抽出部12、32に供給する場合について述べたが、本発明はこれに限らず、例えば図1及び図6との対応部分に同一の符号を付して示す図7及び図8に示すように、ピッチ検出部11'、31'で検出されたピッチ数をピッチクラスデータD11、D31としてクラス分類部14'及び

34' に供給するようにしても良い。

【0078】この場合、クラス分類部14' 及び34' は、可変クラス分類部抽出部12から供給されるピッチクラスデータD11、D31と同様の時間領域に分割されたクラスタップデータD12、D32のクラスコードclass と、ピッチクラスデータD11、D31のクラスコードCLASS とを対応させて統合したクラスコードデータ(class') D14、D34を生成する。

【0079】このようにピッチ検出部11'、31' で検出されたピッチ数をピッチクラスデータD11、D31としてクラス分類部14' 及び34' に供給するようにすれば、クラス分類の頻度を一段と多くすることができ、かくして、クラス分類された結果に基づく予測係数を用いて入力オーディオデータの予測演算を行うオーディオ信号処理装置10は、一段と高音質のオーディオデータに変換することができる。

【0080】また上述の実施の形態においては、オーディオ波形のピッチを検出する方法として自己相関を用いる場合について述べたが、本発明はこれに限らず、例えばオーディオ波形に対してフーリエ変換を施すケプストラム分析やパターン認識等、他の種々の方法を適用することができる。

【0081】また上述の実施の形態においては、入力オーディオデータD10の時間軸領域の波形について、ピッチ検出及びクラス分類を行う場合について述べたが、本発明はこれに限らず、周波数領域に展開してピッチ検出、クラス分類を行うようにしても良い。

【0082】また上述の実施の形態においては、クラス分類のためのタップの切り出しサイズと、予測演算用のタップの切り出し領域とを同一サイズとした場合について述べたが、本発明はこれに限らず、クラス分類用のタップの切り出しサイズと予測演算用のタップの切り出しサイズは異なるようにしても良く、さらには、ピッチ検出部11において検出された自己相関の強さに応じて切り出しサイズを変化させるようにしても良い。この場合、例えば自己相関が強い程切り出しサイズを大きくすることができる。

【0083】また上述の実施の形態においては、予測方式として線形一次による手法を用いる場合について述べたが、本発明はこれに限らず、要は学習した結果を用いるようにすれば良く、例えば多次関数による手法等の種

々の予測方式を適用することができる。

【0084】また上述の実施の形態においては、クラス分類部14においてADRCにより圧縮データパターンを生成する場合について述べたが、本発明はこれに限らず、可逆符号化(DPCM:Differential Pulse Code Modulation) 又はベクトル量子化(VQ:Vector Quantize) 等の圧縮手段を用いるようにしても良い。

【0085】また上述の実施の形態においては、学習回路30の生徒信号生成フィルタ37において教師オーディオデータD30から所定サンプルを間引く場合について述べたが、本発明はこれに限らず、例えばビット数を間引く等、他の種々の方法を適用することができる。

【0086】

【発明の効果】上述のように本発明によれば、デジタルオーディオ信号のピッチを検出し、当該検出したピッチに基づいてそのクラスを分類し、分類されたクラスに対応した予測方式でデジタルオーディオ信号を変換するようにしたことにより、一段とデジタルオーディオ信号の特徴に適応した変換を行うことができる。

【図面の簡単な説明】

【図1】本発明によるデジタル信号処理装置の構成を示すブロック図である。

【図2】デジタルオーディオ信号のピッチの説明に供する信号波形図である。

【図3】ピッチの無いデジタルオーディオ信号の処理の説明に供する信号波形図である。

【図4】オーディオ信号処理装置の構成を示すブロック図である。

【図5】オーディオ信号変換処理手順を示すフローチャートである。

【図6】本発明による学習装置の構成を示すブロック図である。

【図7】他の実施の形態によるデジタル信号処理装置の構成を示すブロック図である。

【図8】他の実施の形態による学習装置の構成を示すブロック図である。

【符号の説明】

10……オーディオ信号処理装置、11、31……ピッチ検出部、14、34……クラス分類部、15……予測係数メモリ、16……予測演算部、36……予測係数算出部、37……生徒信号生成フィルタ。



【図1】

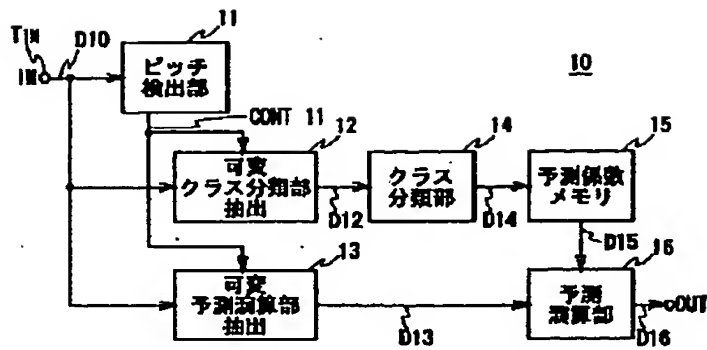


図1 オーディオ信号処理装置の構成

【図2】

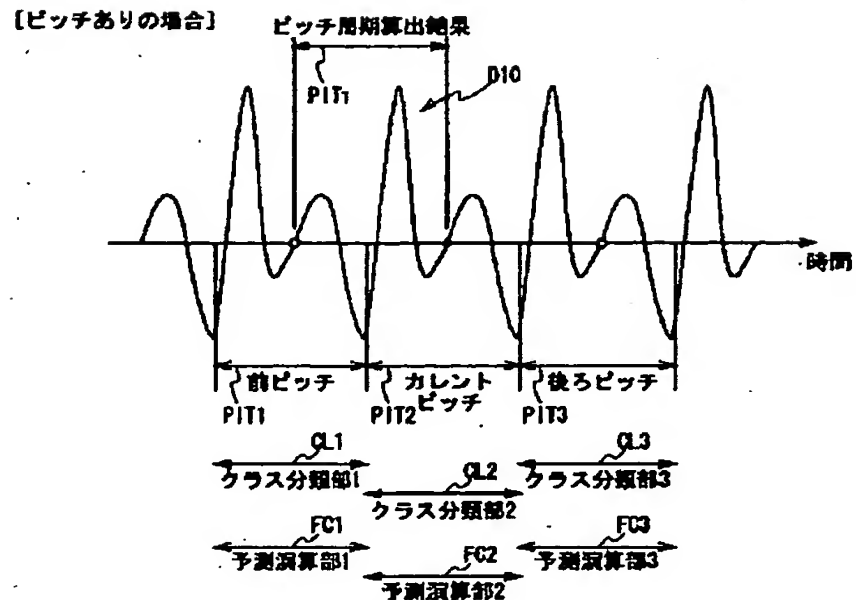


図2 ピッチに基づくクラス分類適応処理

【図3】

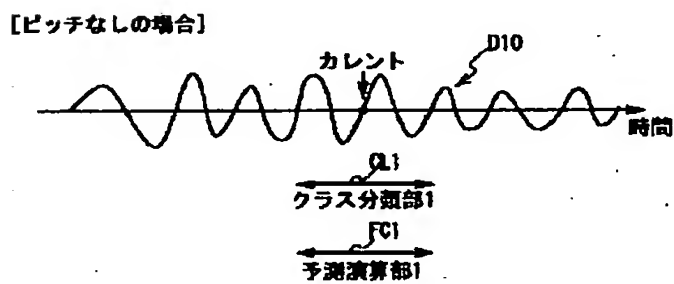


図3 ピッチなしの場合の処理

【図4】

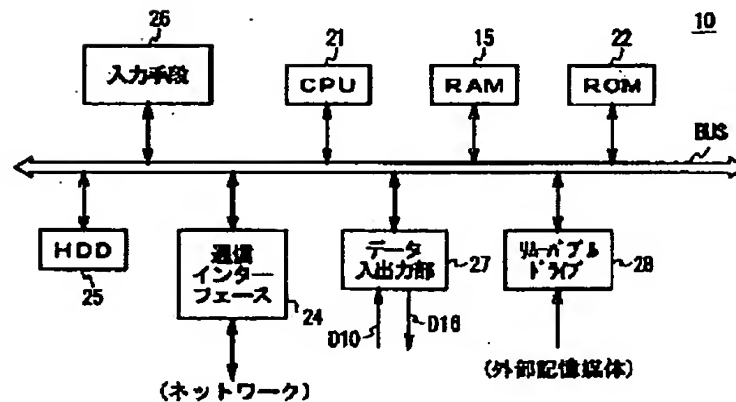


図4 オーディオ信号処理装置

【図5】

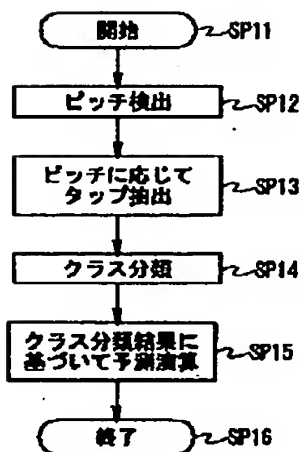


図5 オーディオデータの交換処理手順

【図6】

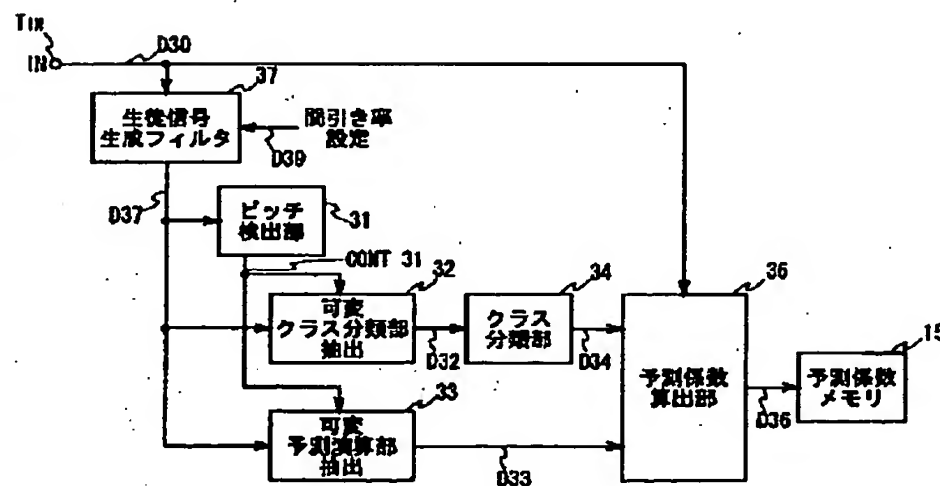


図6 学習回路

【図7】

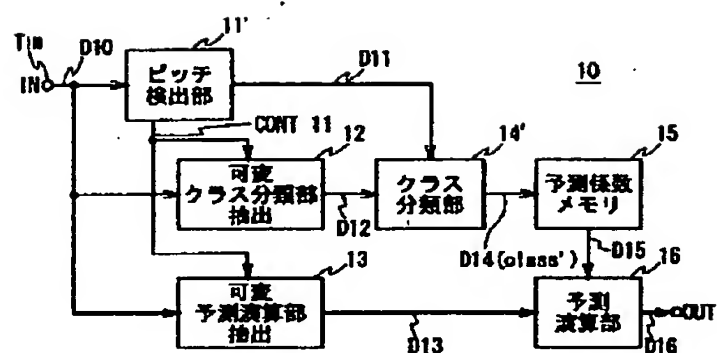


図7 他の実施の形態によるオーディオ信号処理装置の構成

【図8】

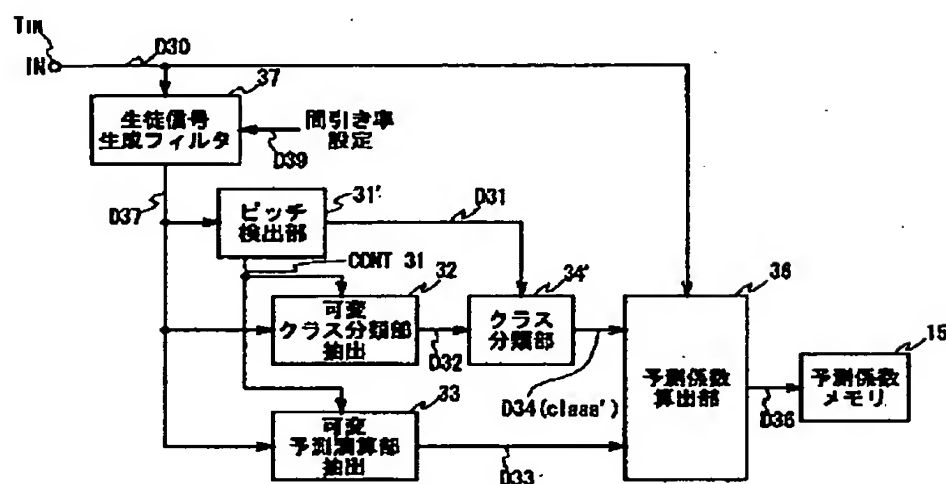


図8 他の実施の形態による学習回路の構成

# PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2002-049395

(43)Date of publication of application : 15.02.2002

(51)Int.Cl.

G10L 19/04  
G10L 11/04  
H03M 7/36  
// G10L101:12

(21)Application number : 2000-238891

(71)Applicant : SONY CORP

(22)Date of filing : 02.08.2000

(72)Inventor : KONDO TETSUJIRO  
WATANABE TSUTOMU

(54) DIGITAL SIGNAL PROCESSING METHOD, LEARNING METHOD, AND THEIR APPARATUS, AND PROGRAM STORAGE MEDIA THEREFOR

(57)Abstract:

PROBLEM TO BE SOLVED: To provide a digital signal processing method further improving the waveform reproducibility of a digital signal, a learning method, and their apparatus, and program storage media therefor.

SOLUTION: In this method, the pitch of a digital audio signal D10 is detected, it is classified based on the detected pitch, and the digital audio signal D10 is converted by a prediction system corresponding to the categorized class, and thereby a conversion further adaptive to a characteristic of the digital audio signal D10 can be performed.

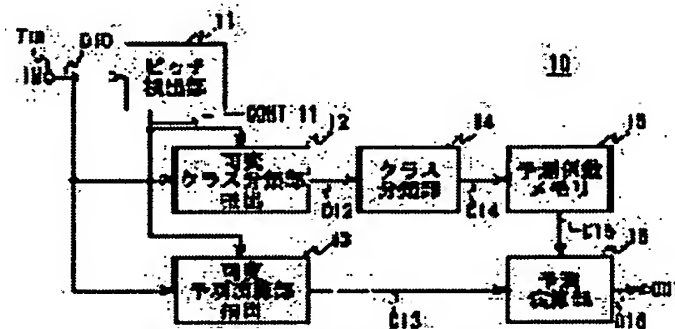


図1 オーディオ信号処理装置の構成

## LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's  
decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

## \* NOTICES \*

Japan Patent Office is not responsible for any damages caused by the use of this translation.

- 1.This document has been translated by computer. So the translation may not reflect the original precisely.
- 2.\*\*\*\* shows the word which can not be translated.
- 3.In the drawings, any words are not translated.

---

**CLAIMS**


---

**[Claim(s)]**

[Claim 1] The digital-signal-processing equipment carry out having a prediction operation means generate the new digital audio signal which comes to change the above-mentioned digital audio signal by carrying out the prediction operation of the above-mentioned digital audio signal in the digital-signal-processing equipment which changes a digital audio signal by pitch detection means detect the pitch of the above-mentioned digital audio signal, class classification means classify the class based on the above-mentioned pitch, and the prediction method corresponding to the class by which the classification was carried out [ above-mentioned ] as the description.

[Claim 2] The above-mentioned prediction operation means is digital-signal-processing equipment according to claim 1 characterized by using the prediction coefficient currently generated by study based on the digital audio signal beforehand considered as a request.

[Claim 3] The digital-signal-processing approach of carrying out having the prediction operation step which generates in the digital-signal-processing approach of changing a digital audio signal, in the pitch detection step which detects the pitch of the above-mentioned digital audio signal, the class classification step into which the class classifies based on the above-mentioned pitch, and the new digital audio signal which come to change the above-mentioned digital audio signal by carrying out the prediction operation of the above-mentioned digital audio signal by the prediction method corresponding to the class by which the classification was carried out [ above-mentioned ] as the description.

[Claim 4] The digital-signal-processing approach according to claim 3 characterized by using the prediction coefficient currently generated by study at the above-mentioned prediction operation step based on the digital audio signal beforehand considered as a request.

[Claim 5] In the study equipment which generates the prediction coefficient used for the prediction operation of the above-mentioned transform processing of the digital-signal-processing equipment which changes a digital audio signal A student digital audio signal generation means to generate the student digital audio signal which degraded the digital audio signal concerned from the digital audio signal considered as a request, A pitch detection means to detect the pitch of the above-mentioned student digital audio signal, Study equipment characterized by having a prediction coefficient calculation means to compute the prediction coefficient corresponding to the above-mentioned class based on a class classification means to classify the class based on the above-mentioned pitch, and the above-mentioned digital audio signal and the above-mentioned student digital audio signal.

[Claim 6] In the study approach which generates the prediction coefficient used for the prediction operation of the above-mentioned transform processing of the digital-signal-processing equipment which changes a digital audio signal The student digital audio signal generation step which generates the student digital audio signal which degraded the digital audio signal concerned from the digital audio signal considered as a request, The pitch detection step which detects the pitch of the above-mentioned student digital audio signal, The study approach characterized by having the prediction coefficient calculation step which computes the prediction coefficient corresponding to the above-mentioned class based on the class classification step into which the class is classified based on the above-mentioned pitch, and the



above-mentioned digital audio signal and the above-mentioned student digital audio signal.

[Claim 7] The program storing medium which makes digital-signal-processing equipment perform the program contain the prediction operation step which generates in the new digital audio signal which comes to change the above-mentioned digital audio signal by carrying out the prediction operation of the above-mentioned digital audio signal by the pitch detection step which detects the pitch of the above-mentioned digital audio signal, the class classification step into which the class is classified based on the above-mentioned pitch, and the prediction method corresponding to the class by which the classification was carried out [ above-mentioned ].

[Claim 8] The program storing medium which makes study equipment perform the program containing the prediction coefficient calculation step which computes the prediction coefficient corresponding to the above-mentioned class based on the student digital audio signal generation step which generates the student digital audio signal which degraded the digital audio signal concerned from the digital audio signal considered as a request, the pitch detection step which detect the pitch of the above-mentioned student digital audio signal, the class classification step into which the class classifies based on the above-mentioned pitch, and the above-mentioned digital audio signal and the above-mentioned student digital audio signal.

---

[Translation done.]

**\* NOTICES \***

Japan Patent Office is not responsible for any damages caused by the use of this translation.

1. This document has been translated by computer. So the translation may not reflect the original precisely.
2. \*\*\*\* shows the word which can not be translated.
3. In the drawings, any words are not translated.

---

**DETAILED DESCRIPTION**

---

**[Detailed Description of the Invention]****[0001]**

**[Field of the Invention]** This invention relates to a program storing medium at the digital-signal-processing approach, the study approaches, and those equipment lists, and is a rate converter or PCM (Pulse Code Modulation). It applies to a program storing medium and is suitable for the digital-signal-processing approach of performing interpolation processing of data to a digital signal in decode equipment etc., the study approaches, and those equipment lists.

**[0002]**

**[Description of the Prior Art]** Before inputting a digital audio signal into digital one/analog converter conventionally, exaggerated sampling processing which changes a sampling frequency by several times the original value is performed. Thereby, the digital audio signal outputted from digital one/analog converter is made as [ eliminate / the phase characteristic of an analog anti-aliasing filter is kept constant in an audio frequency quantity region, and / the effect of the image noise of the digital system accompanying a sampling ].

**[0003]** In this exaggerated sampling processing, the digital filter of a linearity primary (straight line) interpolation method is usually used. Such a digital filter generates linear interpolation data in quest of the average value of two or more existing data, when a sampling rate changes or data are missing.

**[0004]**

**[Problem(s) to be Solved by the Invention]** However, although the amount of data has become precise several times to time amount shaft orientations by linearity linear interpolation as for the digital audio signal after exaggerated sampling processing, the frequency band of the digital audio signal after exaggerated sampling processing seldom changes to before conversion, and the tone quality itself has not improved. Furthermore, since the interpolated data were not necessarily generated based on the wave of the analog audio signal in front of A/D conversion, most its wave repeatability has not improved.

**[0005]** Moreover, although the frequency was changed using the sampling rate converter when the digital audio signal from which a sampling frequency differs was dubbed, it was difficult to be able to perform only interpolation of linear data but to improve tone quality and wave repeatability with a linearity primary digital filter, also by this case. Furthermore, it is the same when the data sample of a digital audio signal is missing.

**[0006]** This invention was made in consideration of the above point, and tends to propose a program storing medium in the digital-signal-processing approach which may improve the wave repeatability of a digital signal much more, the study approaches, and those equipment lists.

**[0007]**

**[Means for Solving the Problem]** Since this technical problem is solved, conversion which was adapted for the description of a digital audio signal much more can be performed by detecting the pitch of a digital audio signal in this invention, classifying the class based on the detected pitch concerned, and having changed the digital audio signal by the prediction method corresponding to the classified class.

**[0008]**

[Embodiment of the Invention] About a drawing, the gestalt of 1 operation of this invention is explained in full detail below.

[0009] In case the audio signal processor 10 raises the sampling rate of a digital audio signal (this is called audio data below) or audio data are interpolated in drawing 1, it is made as [ generate / the audio data near a true value / class classification application processing ].

Incidentally, a digital audio signal means the sound signal showing the voice which people and an animal utter, the musical-sound signal showing the musical sound which a musical instrument emits, and the signal showing other sounds.

[0010] That is, in the audio signal processor 10, the pitch detecting element 11 detects the pitch (fundamental frequency) of the input audio data D10 shown in drawing 2 or drawing 3 supplied from the input terminal TIN based on the autocorrelation of the input audio data D10.

[0011] That is, it is judged that the pitch detecting element 11 detects an autocorrelation value for every range of the versatility which it comes to divide into a time-axis field, and it has a pitch about the input audio data D10 when the autocorrelation value concerned is larger than a predetermined threshold. And the pitch detecting element 11 is the pitch period PITT in case there is a pitch. And let pitch size and its location be pitch classes. For example, when it has the wave as the input audio data D10 show to drawing 2, the pitch detecting element 11 detects front pitch PIT1 and the back pitch PIT 3 in the location which adjoins before and after that to the current pitch PIT 2. Incidentally, front pitch PIT1 and the back pitch PIT 3 which the pitch detecting element 11 detects do not necessarily restrict adjoining the current pitch PIT 2, but also have a predetermined distance detached building \*\*\*\*\* case from the current pitch PIT 2.

[0012] The pitch detecting element 11 will supply the logging control signal CONT11 for starting these to the adjustable class classification section extract section 12 and the adjustable prediction operation part extract section 13, if front pitch PIT1 and the back pitch PIT 3 corresponding to the current pitch PIT 2 are detected.

[0013] Based on the logging control signal CONT11, the adjustable class classification section extract section 12 starts the class tap CL 3 corresponding to the class tap CL 1 corresponding to front pitch PIT1, the class tap CL 2 corresponding to the current pitch PIT 2, and the back tap PIT 3 from the input audio data D10, and supplies it to the class classification section 14 by making these into the class tap data D12.

[0014] The class classification section 14 is ADRC (Adaptive Dynamic Range Coding) which unifies these individually in each pitch (front pitch PIT1, the current pitch PIT 2, and the back tap PIT 3) of every, compresses the class tap, and generates a compression data pattern about the class tap started in the class classification extract section 12. It has the circuit section and the class code generating circuit section which generates the class code to which the class tap data D12 belong.

[0015] The ADRC circuit section forms pattern compressed data by performing an operation which is compressed into 2 bits from 8 bits as opposed to the class tap data D12. Since this ADRC circuit section can perform accommodative quantization and can express the local pattern of signal level efficiently by the short word length here, it is used for code generating of a class classification of a signal pattern.

[0016] When it is going to carry out the class classification of the six 8-bit data on an audio wave (class tap), it must classify into a huge number 248 of classes, and, specifically, the burden on a circuit increases. So, in the class classification section 14 of the gestalt of this operation, a class classification is performed based on the pattern compressed data generated in the ADRC circuit section prepared in that interior. For example, if 1-bit quantization is performed to six class taps, six class taps can be expressed with 6 bits, and can be classified into  $2^6 = 64$  class.

[0017] Here, the ADRC circuit section is a degree type and [0018], when the data level of m and each class tap is set to L and a quantization code is set [ the dynamic range of the class tap in the started field (audio wave) ] to Q for this [ DR and / bit rate ].

[Equation 1]

$$DR = MAX - MIN + 1$$

$$Q = \lceil (L - MIN + 0.5) \times 2^m / DR \rceil \quad \dots\dots (1)$$

[0019] It is alike, and it follows and quantizes by dividing equally by the bit length which had between the maximum MAX in a field, and the minimum values MIN specified. In addition, in (1) type, [ ] means the cut-off processing below decimal point. In this way, supposing six class taps consist of 8 bits ( $m=8$ ), for example, respectively, as for these, each will be compressed into 2 bits in the ADRC circuit section.

[0020] Thus, if the compressed audio data point (class tap) is set to  $q_n$  ( $n=1-6$ ), respectively, the class code generating circuit section prepared in the class classification section 14 is the compressed audio data point  $q_n$ . It is based and is a degree type and [0021].

[Equation 2]

$$\text{class} = \sum_{i=1}^n q_i \cdot (2^P)^{i-1} \quad \dots\dots (2)$$

[0022] Class code class which shows the class to which the block ( $q_1 - q_6$ ) belongs by being alike and performing the shown operation Class code class based on the computed class tap data D12 concerned while computing It considers as the class code data D14, and the prediction coefficient memory 15 is supplied. This class code class The read-out address at the time of reading a prediction coefficient from the prediction coefficient memory 15 is shown. Incidentally it is the audio data point (class tap)  $q_n$  into which  $n$  was compressed in (2) types. A number is expressed and, in the case of the gestalt of this operation, it is  $n=6$ , and  $P$  expresses bit assignment and, in the case of the gestalt of this operation, is  $P=2$ .

[0023] Thus, the class classification section 14 is the class code of the class tap started from the input audio data D10 in the class classification section extract section 12. class is generated and the prediction coefficient memory 15 is supplied by making this into the class code data D14.

[0024] Set  $w_1 - w_n$  of the prediction coefficient which the set of the prediction coefficient corresponding to each class code is memorized to the address corresponding to a class code by the prediction coefficient memory 15, respectively, and is memorized to the address corresponding to the class code concerned based on the class code data D14 supplied from the class classification section 14 It is read and the prediction operation part 16 is supplied.

[0025] The prediction operation part 16 is [ the audio data point (prediction tap) D13 ( $x_1 - x_n$ ) which was started according to the logging control signal CONT11 supplied from the pitch detecting element 11 in the prediction operation part extract section 13 and which is going to carry out a prediction operation, and ] prediction coefficients  $w_1 - w_n$ . It receives and is a degree type [0026].

[Equation 3]

$$y' = w_1 \cdot x_1 + w_2 \cdot x_2 + \dots\dots + w_n \cdot x_n \quad \dots\dots (3)$$

[0027] Prediction result  $y'$  is obtained by being alike and performing the shown sum-of-products operation. This forecast  $y'$  is outputted from the prediction operation part 16 as audio data D16 with which tone quality has been improved.

[0028] When it is incidentally judged that there is no pitch in the input audio data D10 in the pitch detecting element 11, the pitch detecting element 11 performs a class classification and a prediction operation by the audio wave near the current data by supplying the control signal CONT11 for starting the level of an audio wave near the current data, as shown in drawing 3 to the adjustable class classification section extract section 12 and the adjustable prediction operation part extract section 13, without using front pitch PIT1 and the back pitch PIT 3.

[0029] In addition, although functional block mentioned above about drawing 1 as a configuration of the audio signal processor 10 was shown, the equipment of a computer configuration shown in drawing 4 in the gestalt of this operation as a concrete configuration which constitutes this functional block is used. In drawing 4 namely, the audio signal processor 10 It has RAM (Random Access Memory)15 which constitutes CPU21, ROM (Read Only Memory)22, and the prediction coefficient memory 15 through Bus BUS, and the configuration to which each circuit section was connected, respectively. By performing the various programs stored in ROM22, CPU11 It is made

as [ operate / as each functional block (the pitch detecting element 11, the class classification section extract section 12, the prediction operation part extract section 13, the class classification section 14, and prediction operation part 16) mentioned above about drawing 1 ].

[0030] Moreover, it has the removable drive 28 which reads information from external storage, such as the communication link interface 24 and floppy disk which communicate between networks, and a magneto-optic disk, to the audio signal processor 10, each program for performing class classification application processing mentioned above about drawing 1 from a network course or external storage can be read to the hard disk of a hard disk drive unit 25, and class classification adaptation processing can also be performed according to \*\*\*\*\* and the read program concerned.

[0031] A user performs class classification processing mentioned above about drawing 1 to CPU21 by inputting various commands through the input means 26, such as a keyboard and a mouse. In this case, after the audio signal processor 10 inputs the audio data (input audio data) D10 which are going to raise tone quality through the data I/O section 27 and performs class classification application processing to the input audio data D10 concerned, it is made as [ output / through the data I/O section 27 / the audio data D16 whose tone quality improved / outside ].

[0032] Incidentally, drawing 5 shows the procedure of the class classification adaptation processing in the audio signal processor 10, and if the audio signal processor 10 goes into the procedure concerned from a step SP 11, it will compute the pitch of the input audio data D10 in the pitch detecting element 11 in the continuing step SP 12.

[0033] This computed pitch (front pitch PIT1, the current pitch PIT 2, the back pitch PIT 3) is also that for ensuring the class classification of an OTIO wave much more, and the audio signal processor 10 carries out the class classification of the class tap (audio wave) by the class classification section 14 in a step SP 14, after starting the class tap according to a pitch in a step SP 13. And the audio signal processor 10 reads a prediction coefficient from the prediction coefficient memory 15 using the class code obtained as a result of the class classification. This prediction coefficient corresponds for every class by study beforehand, and is stored, and the audio signal processor 10 can use the prediction coefficient corresponding to the description of an audio wave at this time by reading the prediction coefficient corresponding to a class code.

[0034] The prediction coefficient read from the prediction coefficient memory 15 is used for the prediction operation of the prediction operation part 16 in a step SP 15. Thereby, the input audio data D10 are changed into the audio data D16 which a prediction operation is carried out using the class code obtained by the class tap according to the pitch, and are considered as a request. In this way, the input audio data D10 are changed into the audio data D16 with which the tone quality has been improved, and the audio signal processor 10 moves to a step SP 16, and ends the procedure concerned.

[0035] Next, the study circuit for obtaining beforehand the set of the prediction coefficient for every class memorized to the prediction coefficient memory 15 mentioned above about drawing 1 by study is explained.

[0036] In drawing 6, the study circuit 30 receives the teacher audio data D30 of the quality of loud sound in the student signal generation filter 37. The student signal generation filter 37 is made as [ lengthen / the teacher audio data D30 / for every predetermined time / at the rate of infanticide set up by the rate setting signal D39 of infanticide / between predetermined samples ].

[0037] In this case, the prediction coefficient generated changes with rates of infanticide in the student signal generation filter 37, and the audio data reproduced with the above-mentioned audio signal processor 10 according to this also differ. For example, when it is going to improve the tone quality of audio data by making a sampling frequency high in the above-mentioned audio signal processor 10, infanticide processing which reduces a sampling frequency is performed with the student signal generation filter 37. Moreover, when aiming at improvement in tone quality by compensating the data sample which lacked the input audio data D10 in the above-mentioned audio signal processor 10 to this, according to this, it is made as [ perform / infanticide processing made to lack a data sample ] with the student signal generation filter 37.



[0038] In this way, the student signal generation filter 37 generates the student audio data D37 by predetermined infanticide processing from the teacher audio data 30, and supplies this to the pitch detecting element 31, the class classification section extract section 32, and the prediction operation part extract section 33, respectively.

[0039] About the student audio data D37 supplied from the student signal generation filter 37, the pitch detecting element 31 detects the pitch (fundamental frequency) mentioned above about drawing 2 and drawing 3 based on the autocorrelation of the student audio data D37.

[0040] And the pitch detecting element 31 will supply the logging control signal CONT31 for starting these to the adjustable class classification section extract section 32 and the adjustable prediction operation part extract section 33, if front pitch PIT1 and the back pitch PIT 3 corresponding to the current pitch PIT 2 are detected.

[0041] Based on the logging control signal CONT31, the adjustable class classification section extract section 32 starts the class tap CL 3 corresponding to the class tap CL 1 corresponding to front pitch PIT1, the class tap CL 2 corresponding to the current pitch PIT 2, and the back tap PIT 3 from the student audio data D37, and supplies it to the class classification section 34 by making these into the class tap data D32.

[0042] The class classification section 34 is ADRC (Adaptive Dynamic Range Coding) which compresses the class tap concerned and generates a compression data pattern about the class tap started in the class classification extract section 32. It has the circuit section and the class code generating circuit section which generates the class code to which the class tap data D32 belong.

[0043] The ADRC circuit section forms pattern compressed data by performing an operation which is compressed into 2 bits from 8 bits as opposed to the class tap data D32. Since this ADRC circuit section can perform accommodative quantization and can express the local pattern of signal level efficiently by the short word length here, it is used for code generating of a class classification of a signal pattern.

[0044] When it is going to carry out the class classification of the six 8-bit data on an envelope wave (class tap), it must classify into a huge number 248 of classes, and, specifically, the burden on a circuit increases. So, in the class classification section 14 of the gestalt of this operation, a class classification is performed based on the pattern compressed data generated in the ADRC circuit section prepared in that interior. For example, if 1-bit quantization is performed to six class taps, six class taps can be expressed with 6 bits, and can be classified into  $2^6 = 64$  class.

[0045] Here, the ADRC circuit section sets [ the dynamic range of the class tap in the started field (audio wave) ] a quantization code to Q for this [ DR and / bit rate ], setting the data level of m and each class tap as L, and quantizes by dividing equally by the bit length which had between the maximum MAX in a field, and the minimum values MIN specified by the same operation as above-mentioned (1) type. In this way, supposing six data points on an audio wave consist of 8 bits ( $m = 8$ ), for example, respectively, as for these, each will be compressed into 2 bits in the ADRC circuit section.

[0046] If the compressed audio data point (class tap) is set to  $q_n$  ( $n=1-6$ ), respectively, thus, the class code generating circuit section prepared in the class classification section 34 Compressed audio data point  $q_n$  By being based and performing the same operation as above-mentioned (2) types Class code class which shows the class to which the block ( $q_1 - q_6$ ) belongs It computes and is the computed class code class concerned. It considers as the class code data D34, and the prediction coefficient calculation section 36 is supplied. Incidentally it is the audio data point (class tap)  $q_n$  into which n was compressed in (2) types. A number is expressed and, in the case of the gestalt of this operation, it is  $n = 6$ , and P expresses bit assignment and, in the case of the gestalt of this operation, is  $P = 2$ .

[0047] Thus, the class classification section 34 generates the class code data D34, and supplies this to the prediction coefficient calculation section 36. Moreover, the audio data point (prediction tap) D13 ( $x_1 - x_n$ ) which was started according to the logging control signal CONT11 supplied from the pitch detecting element 11 in the prediction operation part extract section 33 and which is going to carry out a prediction operation is supplied to the prediction coefficient calculation section 36.

[0048] the class code class and each class code class to which the prediction coefficient calculation section 36 was supplied from the class classification section 34 every -- a normal equation is stood using the started prediction tap D33 and the teacher audio data D30 of the quality of loud sound supplied from the input edge TIN.

[0049] namely, the level of n sample of the student audio data D37 -- respectively --  $x_1, x_2, \dots, x_n$  \*\*\*\*\* -- the quantization data of the result of having been alike, respectively and having performed p-bit ADRC --  $q_1, \dots, q_n$  \*\* -- it carries out. At this time, class code class' of this field is defined like above-mentioned (2) types. And they are  $x_1, x_2, \dots, x_n$  about the level of the student audio data D37 as mentioned above, respectively. When it carries out and level of the teacher audio data D30 of the quality of loud sound is set to y, they are a prediction coefficient  $w_1$ , and  $w_2, \dots, w_n$  for every class code. The linearity presumption type of n tap to depend is set up. It is this A degree type and [0050]

[Equation 4]

$$y = w_1 x_1 + w_2 x_2 + \dots + w_n x_n \quad \dots (4)$$

[0051] It carries out. Before study, it is  $w_n$ . It is an undetermined coefficient.

[0052] In the study circuit 30, it learns to two or more audio data for every class code. When a data measurement size is M, above-mentioned (4) types are followed, and it is a degree type and [0053].

[Equation 5]

$$y_k = w_1 x_{k1} + w_2 x_{k2} + \dots + w_n x_{kn} \quad \dots (5)$$

[0054] It \*\*\*\*\* however,  $k = 1, 2, \dots$  it is M.

[0055] In  $M > n$ , they are a prediction coefficient  $w_1$  and  $\dots w_n$ . Since it is not decided uniquely, it is the element of the error vector e A degree type and [0056]

[Equation 6]

$$e_k = y_k - \{w_1 x_{k1} + w_2 x_{k2} + \dots + w_n x_{kn}\} \quad \dots (6)$$

[0057] It is defined as alike (however,  $k = 1, 2, \dots, M$ ), and is a degree type and [0058].

[Equation 7]

$$e^2 = \sum_{k=1}^M e_k^2 \quad \dots (7)$$

[0059] It asks for the prediction coefficient made into min. It is a solution method by the so-called least square method.

[0060]  $w_n$  according to (7) types here It asks for a partial differential coefficient. In this case, a degree type, [0061]

[Equation 8]

$$\begin{aligned} \frac{\partial e^2}{\partial w_i} &= \sum_{k=1}^M 2 \left( \frac{\partial e_k}{\partial w_i} \right) e_k = \sum_{k=1}^M 2 x_{ki} \cdot e_k \\ &= \sum_{k=1}^M 2 x_{ki} \cdot e_k \quad (i = 1, 2, \dots, n) \quad \dots (8) \end{aligned}$$

[0062] What is necessary is just to calculate each  $w_n$  ( $n=1-6$ ) so that it may be made "0."

[0063] And a degree type, [0064]

[Equation 9]

$$X_{ij} = \sum_{p=1}^M x_{pi} \cdot x_{pj} \quad \dots (9)$$

[0065]

[Equation 10]

$$Y_i = \sum_{k=1}^M x_{ki} \cdot y_k \quad \dots (10)$$

[0066] \*\* -- like --  $X_{ij}$  and  $Y_i$  if a definition is given -- (8) types -- a matrix -- using -- a

degree type and [0067]

[Equation 11]

$$\begin{bmatrix} X_{11} & X_{12} & \dots & X_{1n} \\ X_{21} & X_{22} & \dots & X_{2n} \\ \vdots & \vdots & & \vdots \\ X_{n1} & X_{n2} & \dots & X_{nn} \end{bmatrix} \begin{bmatrix} W_1 \\ W_2 \\ \vdots \\ W_n \end{bmatrix} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \quad \dots (11)$$

[0068] It is expressed by carrying out.

[0069] Generally this equation is called the normal equation. In addition, it is  $n=6$  here.

[0070] the prediction coefficient calculation section 36 after the input of all the data for study (the teacher audio data D30, the class code class, and prediction tap D33) is completed -- each class code class the normal equation having shown in above-mentioned (11) equations -- standing -- this normal equation -- sweeping out -- general matrix solution methods, such as law, -- using -- every --  $W_n$  \*\*\*\*\* -- it solves and a prediction coefficient is computed for every class code. The prediction coefficient calculation section 36 writes each computed prediction coefficient (D36) in the prediction coefficient memory 15.

[0071] As a result of performing such study, in the prediction coefficient memory 15, they are the quantization data  $q_1, \dots, q_6$ . The prediction coefficient for presuming the audio data  $y$  of the quality of loud sound is stored for every class code for every pattern specified. This prediction coefficient memory 15 is used in the audio signal processor 10 mentioned above about drawing 1. By this processing, study of the prediction coefficient for creating the audio data of the quality of loud sound from the usual audio data according to a linearity presumption type is completed.

[0072] Thus, the study circuit 30 can generate the prediction coefficient for the interpolation processing in the audio signal processor 10 in consideration of extent which performs interpolation processing in the audio signal processor 10 by performing infanticide processing of the teacher audio data of the quality of loud sound with the student signal generation filter 37.

[0073] In the above configuration, the audio signal processor 10 can extract the wave part which had especially regularity among the input audio data D10 by detecting the pitch of the input audio data D10 in the pitch detecting element 11.

[0074] When they differ for every individual even if the wave part (pitch) with this regularity is the phoneme same in human being's voice, starts this part correctly and carries out a class classification, the class division of the different description of an audio wave for every individual is carried out clearly much more.

[0075] Thus, by performing the prediction operation based on a class part opium poppy and the class classification result concerned for the description of an audio wave clearly, the repeatability of an audio wave improves and improvement in tone quality is achieved.

[0076] According to the above configuration, the input audio data D10 are convertible for the audio data D16 of the quality of loud sound much more by having controlled the logging range of a tap based on the detection result of a pitch.

[0077] In addition, although the case where the adjustable class classification section extract sections 12 and 32 were supplied in the audio signal processor 10 and the study circuit 30 in the gestalt of above-mentioned operation by making into a pitch class the pitch size detected by the pitch detecting elements 11 and 31 and its location was described As this invention is shown in drawing 7 and drawing 8 which attach and show the same sign to a corresponding point not only with this but drawing 1 and drawing 6 You may make it supply the number of pitches detected by pitch detecting-element 11' and 31' to class classification section 14' and 34' as pitch class data D11 and D31.

[0078] In this case, class classification section 14' and 34' are the class code class of the pitch class data D11 and D31 supplied from the adjustable class classification section extract section 12, and the class tap data D12 and D32 divided into the same time domain. Class code CLASS of the pitch class data D11 and D31 The class code data (class ') D14 and D34 which were made to

correspond and were unified are generated.

[0079] Thus, if the number of pitches detected by pitch detecting-element 11' and 31' is supplied to class classification section 14' and 34' as pitch class data D11 and D31 The frequency of a class classification can be made [ many ] much more, and the audio signal processor 10 which performs the prediction operation of input audio data in this way using the prediction coefficient based on the result by which the class classification was carried out can be changed into the audio data of the quality of loud sound much more.

[0080] Moreover, in the gestalt of above-mentioned operation, although the case where an autocorrelation was used as an approach of detecting the pitch of an audio wave was described, this invention can apply other various approaches, such as cepstrum analysis, pattern recognition, etc. which perform the Fourier transform for example, not only to this but to an audio wave.

[0081] Moreover, in the gestalt of above-mentioned operation, although the case where pitch detection and a class classification were performed about the wave of the time-axis field of the input audio data D10 was described, it develops not only to this but to a frequency domain, and this invention may be made to perform pitch detection and a class classification.

[0082] Moreover, although the case where the logging size of the tap for a class classification and the logging field of the tap for a prediction operation were made into the same size was described in the gestalt of above-mentioned operation You may make it the logging size of not only this but the tap for a class classification differ from the logging size of the tap for a prediction operation, this invention is further started according to the strength of the autocorrelation detected in the pitch detecting element 11, and you may make it change size. In this case, it can start, so that an autocorrelation is strong, and size can be enlarged.

[0083] moreover, the result of in short having learned this invention not only in this although the case where the technique by linearity primary was used as a prediction method was described in the gestalt of above-mentioned operation -- using -- it makes -- \*\*\*\*ing -- for example, -- many -- various prediction methods, such as technique by degree function, are applicable.

[0084] moreover -- although the case where ADRC generated a compression data pattern in the class classification section 14 was described in the gestalt of above-mentioned operation -- this invention -- not only this but reversible coding (DPCM:Differential Pulse Code Modulation) Or vector quantization (VQ:Vector Quantize) etc. -- you may make it use a compression means

[0085] Moreover, in the gestalt of above-mentioned operation, although the case where a predetermined sample was thinned out from the teacher audio data D30 in the student signal generation filter 37 of the study circuit 30 was described, this invention can apply other various approaches, such as thinning out not only this but the number of bits.

[0086]

[Effect of the Invention] As mentioned above, according to this invention, the pitch of a digital audio signal can be detected, the class can be classified based on the detected pitch concerned, and conversion which was adapted for the description of a digital audio signal much more can be performed by having changed the digital audio signal by the prediction method corresponding to the classified class.

---

[Translation done.]

**\* NOTICES \***

Japan Patent Office is not responsible for any damages caused by the use of this translation.

- 1.This document has been translated by computer. So the translation may not reflect the original precisely.
- 2.\*\*\*\* shows the word which can not be translated.
- 3.In the drawings, any words are not translated.

---

**DESCRIPTION OF DRAWINGS**

---

**[Brief Description of the Drawings]**

**[Drawing 1]** It is the block diagram showing the configuration of the digital-signal-processing equipment by this invention.

**[Drawing 2]** It is the signal waveform diagram with which explanation of the pitch of a digital audio signal is presented.

**[Drawing 3]** It is the signal waveform diagram with which explanation of processing of a digital audio signal without a pitch is presented.

**[Drawing 4]** It is the block diagram showing the configuration of an audio signal processor.

**[Drawing 5]** It is the flow chart which shows an audio signal transform-processing procedure.

**[Drawing 6]** It is the block diagram showing the configuration of the study equipment by this invention.

**[Drawing 7]** It is the block diagram showing the configuration of the digital-signal-processing equipment by the gestalt of other operations.

**[Drawing 8]** It is the block diagram showing the configuration of the study equipment by the gestalt of other operations.

**[Description of Notations]**

10 [ .. A prediction coefficient memory, 16 / .. Prediction operation part, 36 / .. The prediction coefficient calculation section, 37 / .. Student signal generation filter. ] .... 11 An audio signal processor, 31 .. 14 A pitch detecting element, 34 .. The class classification section, 15

---

**[Translation done.]**